NONLOCAL OPTIMIZATION METHODS BASED ON POTENTIAL THEORY

A. I. Kaplinskii and A. I. Propoi

UDC 517.977

This paper reviews a new approach to nonlocal optimization methods based on replacing the problem of minimizing a function of several variables by an equivalent randomized problem. We examine the variation of the functional of the randomized problem and show that the original problem is reduced to maximizing a potential function which is a solution of a boundary value problem in mathematical physics. We consider methods of the first and the second order (analogs of the gradient descent and Newton methods) that require only the values of the original function for their implementation. Finally, we analyze properties of the potential function and reveal the major structural features of the nonlocal search: the unstable component in the nonperspective search domain, and the operations of space reflection and dilatation in nonperspective and perspective directions, respectively.

1. INTRODUCTION

Solution methods for extremal problems are rather well developed at the present time. However, these methods are generally aimed at the solution of "good" extremal problems (convex, smooth, deterministic) and are based on a local information (such as "gradient at a point").

The principal feature of the "bad" optimization problems (nonconvex, nonsmooth, nondeterministic) is that they require nonlocal information in the process of their solution.

The general way to introduce the nonlocality is randomization of the original problem. For nonlinear programming problems this was done in [1 – 3]. The solution methods of such problems were based on a stochastic modification of the conventional gradient and led to probabilistic iterative methods (see, for example, [4 – 6]).

The aim of this paper is to consider an alternative (dual) approach to the notion of the gradient in the nonsmooth optimization and new methods following from this notion. The approach has deep roots with adaptation and learning methods [7], potential functions theory [8], mathematical systems theory [9] and has been developed during the last decade [10–17].

In this paper we describe in a systematic way the major theoretical results of this approach. We consider the first variation of the randomized problem and give conditions the mean vector field must satisfy in order to improve the current solution. The partition of this vector field into potential and "nondivergent" components reduces the original problem to the maximization of a potential function (which is a solution of a boundary value problem in mathematical physics). We consider methods of the first and the second order (analogs of the gradient descent and Newton methods in application to the potential function) that require for implementation only the values of the original objective function and show how to take into account a priori information about the problem (differentiability of the objective function, hypotheses on the solution location). We analyze the properties of the potential function and reveal the major structural features of the nonlocal search — unstable component in nonperspective domain, the operations of space dilatation and reflection in perspective and nonperspective search directions.

2. PROBLEM RANDOMIZATION

We consider the unconstrained optimization problem

$$f(x) \to \min, \tag{1}$$

where the function $f : R^n \to R$.

In developing algorithms for solution (1) we will have in mind "bad" functions and shall assume that in general only the function values can be measured. However, we would like to have a technology of optimization algorithms

Voronezh State University, Voronezh, Russia. Institute for Systems Analysis, Russian Academy of Sciences, Moscow, Russia. Translated from *Automatika i Telemekhanika*, No. 7, pp. 55–65, 1993. Original article submitted June 25 1992.

0005-1179/93/5407-1077$12.50

design that takes into account an increasing level of information about the problem to be solved (knowledg derivatives, assumptions or constraints on the solution location, etc.).

For this purpose we restate the problem (1), introducing the probability distribution of x, that is, randomiz the problem (1):

$$F(X) = E[f(X)] \to min_{X \in \{X\}},$$

where $\{X\}$ is a set of admissible random vectors with values in R^n. The problem (2) can be also written as

$$\int_{R^n} f(x)p(x)\,dx \to min_{p(x)},$$

where $p(x)$ is the probability density function of the random vector X, understandable, generally speaking, in generalized sense.

The restatement (2) of the problem (1) is based on the following propositions:

Proposition 1. Let the minimum of the problem (1) be achieved and let the set of random vectors $\{X\}$ of problem (2) contain random vectors with discrete probability distributions, i.e., the set of feasible distributions p contains δ-functions.

Then

$$min_{x \in R^n} f(x) = min_{X \in \{X\}} E[f(X)].$$

Proposition 2. Let X^* with the density function $p^*(x)$ be a solution of the problem (2). Then any point $x^* \in$ for which $f(x)$ is continuous and $\int_{\|x-x^*\| \le \epsilon} p^*(x)\,dx > 0$ within $\|x - x^*\| \le \epsilon$ will be a solution of the problem 1).

The proof of these propositions is simple and can be found, for example, in [11].

Consider a nonlinear programming problem

$$\omega = \{\min f(x) | f_j(x) \le 0, j = 1, \ldots, m, x \in R^n\}$$

and its randomized version

$$\bar{\omega} = \{\min \int_{R^n} f(x)p(x)\,dx | \int_{R^n} f_j(x)p(x)\,dx \le 0, j = 1, \ldots, m\}.$$

In this case, unlike (4), $\bar{\omega} \le \omega$, because the set of admissible solutions for problem (6) is generally larger than f the problem (5) (the constraints in (6) may be fulfilled on the average). If (5) is the convex programming proble than statements similar to propositions 1 and 2 are true.

A variant of the problem (2) is

$$\bar{\omega}_\Omega = \{\min \int_\Omega f(x)p(x)\,dx | x \in \Omega\},$$

where Ω is a set in R^n, $\int_\Omega p(x)\,dx = 1$ (for example, this set can describe a priori information about the locati of the problem solution). In this case the statements on problem equivalence also hold even for nonconvex sets (because in this case only the objective function is randomized).

3. VARIATION OF THE RANDOMIZED FUNCTIONAL

We shall consider iterative procedures

$$X_N = X_N + \epsilon_N Y_N, N = 0, 1, 2, \ldots,$$

for solution of the problem (2). Here Y_N is the random direction of changing X_N with a joint probability densit function $p_N(x, y)$, and ϵ_N is the step length along Y_N.

To determine Y_N we compute $\delta_Y F(X)$— the derivative of the functional (2) at the state X along the directio Y.

By definition

$$\delta_Y F(X) = \lim_{\epsilon \to 0} \frac{F(X + \epsilon Y) - F(X)}{\epsilon} =$$

$$\frac{d}{d\epsilon}[F(X + \epsilon Y)]_{\epsilon=0} = \frac{d}{d\epsilon}[F(X_\epsilon)]_{\epsilon=0} =$$

$$[\frac{d}{d\epsilon} \int_{R^n} f(x)p_\epsilon(x)\,dx]_{\epsilon=0},$$

where the density function $p_\epsilon(x)$ of the random vector $X_\epsilon = X + \epsilon Y$ may be expressed by the joint density

$$p_\epsilon(x) = \int_{R^n} p(x - \epsilon y, y)\, dy. \tag{10}$$

Assuming that the conventional conditions for differentiability of integrals with respect to a parameter hold (these conditions can always be satisfied by an appropriate choice of distributions), we have

$$\delta_Y F(X) = [\frac{d}{d\epsilon} \int_{R^n} f(x) \int_{R^n} p(x - \epsilon y, y)\, dy\, dx]_{\epsilon=0} = \int_{R^n} \frac{d}{d\epsilon}[A_\epsilon f(y)]_{\epsilon=0}\, dy. \tag{11}$$

Here

$$A_\epsilon f = \int_{R^n} f(x) p(x - \epsilon y, y)\, dx \tag{12}$$

is a pseudodifferentiable operator [18].

For computing (11) we use the known relation of vector analysis

$$[\frac{d}{d\epsilon} p(x - \epsilon y, y]_{\epsilon=0} = -(\frac{\partial p(x,y)}{\partial x}, y) = -\mathrm{div}_x[p(x,y)y],$$

where $(,)$ is the scalar product in R^n, $\mathrm{div} f = \sum_{i=1}^n \frac{\partial f}{\partial x_i}$.

In result,

$$\delta_Y F(X) = -\int_{R^n} \int_{R^n} f(x)(\frac{\partial p(x,y)}{\partial x}, y)\, dy\, dx$$

$$= -\int_{R^n} \int_{R^n} f(x)\mathrm{div}_x[p(x,y)y]\, dy\, dx = -\int_{R^n} f(x)\mathrm{div}_x[\int_{R^n} p(x,y)y\, dy]\, dx.$$

Introducing the conditional probability density function $p(x,y) = p(x)p(y|x)$ and denoting

$$\bar{y}(x) = \int_{R^n} yp(y|x)\, dy = E[Y|X = x], \tag{13}$$

we can rewrite the above expression as

$$\delta_Y F(X) = -\int_{R^n} f(x)\mathrm{div}[p(x)\bar{y}(x)]\, dx.$$

Differentiating the identity

$$\int_{R^n} p_\epsilon(x)\, dx \equiv 1$$

by ϵ, we find, that

$$\int_{R^n} \mathrm{div}[p(x)\bar{y}(x)]\, dx = 0. \tag{14}$$

Finally, we obtain for the derivative of F along Y the expression

$$\delta_Y F(X) = -\int_{R^n} [f(x) - c]\mathrm{div}[p(x)\bar{y}(x)]\, dx, \tag{15}$$

where c is an arbitrary constant.

The direction Y that leads to a decrease in the functional $F(X)$ must satisfy the inequality $\delta_Y F(X) \leq 0$. In particular, this inequality holds if

$$\mathrm{div}[p(x)\bar{y}(x)] = (f(x) - c)p_U(x); p_U(x) \geq 0.$$

This condition is discussed in details in the next Section.

Let the function $f(x)$ have continuous derivatives of the first order. Then $\delta_Y F(X)$ may be expressed in the following way:

$$\delta_Y F(X) = [\frac{d}{d\epsilon} \int_{R^n} \int_{R^n} f(x + \epsilon y)p(x,y)\, dx\, dy]_{\epsilon=0} =$$

$$\int_{R^n}\int_{R^n}\left(\frac{\partial f(x)}{\partial x}, y\right)p(x,y)\,dx\,dy = \int_{R^n}\left(\frac{\partial f(x)}{\partial x}, \bar{y}(x)\right)p(x)\,dx,$$

where $\bar{y}(x)$ is defined by (13).

Hence, the directional derivative has two equavalent representations:

$$\delta_Y F(X) = \int_{R^n}\left(\frac{\partial f(x)}{\partial x}, \bar{y}(x)\right)p(x)\,dx = -\int_{R^n}(f(x)-c)\mathrm{div}[p(x)\bar{y}(x)]\,dx. \qquad (16$$

The first representation in (16) leads to the known stochastic iterative methods (such as the stochastic approximati method) [7], while the second representation in (16) is a base for new optimization methods, which will be discuss in the following sections.

4. THE NOTION OF GRADIENT IN NONSMOOTH OPTIMIZATION

Now we define the direction Y that maximizes $\delta_Y F(X)$ (under some natural constraints on Y in order to elimina infinite solutions). We will call such a vector the gradient direction. For this purpose we use the Schwarz inequali with a weight function $p_U(x)$:

$$|\delta_Y F(X)| \le \left(\int_{R^n}(f(x)-c)^2 p_U(x)\,dx\right)^{1/2}\left(\int_{R^n}\left\{\frac{\mathrm{div}[p(x)\bar{y}(x)]}{p_U(x)}\right\}^2 p_U(x)\,dx\right)^{1/2}, \qquad (17$$

where the function $p_U(x) > 0$ may be considered as a probability density function of a random vector U. Th appropriate choice of $p_U(x)$ guarantees the existence of the integrals in (17). From (17) it follows that maximizatio of $\delta_Y F(X)$ over Y leads to the equation

$$\frac{\mathrm{div}[p(x)\bar{y}(x)]}{p_U(x)} = \lambda(f(x)-c), \lambda > 0,$$

or

$$\mathrm{div}[p(x)\bar{y}(x)] = \lambda(f(x)-c)p_U(x), \qquad (18$$

where the constant c is defined now by the equality

$$c = \int_{R^n} f(x)p_U(x)\,dx, \qquad (19$$

because of (14).

If we limit the "length" of the vector Y by the equality

$$\int_{R^n}\left\{\frac{\mathrm{div}[p(x)\bar{y}(x)]}{p_U(x)}\right\}^2 p_U(x)\,dx = \alpha^2, \qquad (20$$

where $\alpha > 0$ is a given constant, then the parameter λ in (18) is defined as

$$\lambda = \alpha\left(\int_{R^n}(f(x)-c)^2 p_U(x)\,dx\right)^{-1/2}. \qquad (21$$

The corresponding value of $\delta_Y F(X)$ is

$$\delta_Y F(X) = \alpha\left(\int_{R^n}(f(x)-c)^2 p_U(x)\,dx\right)^{1/2}. \qquad (22$$

If the function $p_U(x)$ is interpreted as a probability density function of a random vector U, then (22) and (19) can be rewritten as

$$c = E[f(U)], \delta_Y F(X) = \alpha\{\sigma^2[f(U)]\}^{1/2}, \qquad (23$$

where $\sigma^2[f(U)]$ is the variance of $f(U)$.

Let us summarize the results obtained.

Proposition 3. The gradient direction Y defined by the mean vector field (13) that maximizes $\delta_Y F(X)$ subject to the constraints (14) and (20) must satisfy the equations (18) and (19). The corresponding value of $\delta_Y F(X)$ for this direction is determined by (22) or (23).

The gradient direction is defined up to the arbitrary function $p_U(x) > 0$. Concentrating $p_U(x)$ near some point x_0: $p_U(x) \to \delta(x - x_0)$, we can reduce nonlocality of the search: $\delta_Y F(X) \to 0$ in (23). On the contrary, concentrating $p_U(x)$ both at the minimum and the maximum points of $f(x)$, we can increase the variance (23), and thus the nonlocality of the search.

One can see the similarity of the above constructions to the derivation of the Kramer-Rao inequality [19]. In particular, the integral in (20) may be associated with the Fisher information function and be interpreted as a measure of the information yielded by the random vector U (which, according to the above remark, defines the nonlocality of the search).

5. POTENTIAL FIELD

The given notion of the gradient vector field is still not constructive, because the equation (18) has an infinity of solutions. Therefore, we partition the vector field $p(x)\bar{y}(x)$ in (18) into the sum of the potential and the nondivergent fields:

$$p(x)\bar{y}(x) = \nabla\varphi(x) + w(x), \tag{24}$$

$$\operatorname{div} w(x) = 0. \tag{25}$$

We assume that the vector field $p(x)\bar{y}(x)$ is finite or vanishes sufficiently rapidly at infinity. Under some additional assumptions, such as $\varphi|_\infty = 0$, the representation (24), (25) is unique.

Taking into account that div grad $\varphi = \Delta\varphi$, where Δ is the Laplace operator, we obtain from (18), (24) and (25)

$$\Delta\varphi(x) = -\lambda[f(x) - c]p_U(x). \tag{26}$$

Thus, the potential function that characterizes the gradient direction should satisfy the Poisson equation (26). Its solution, vanishing at infinity (it always can be assured by the appropriate choice of $p_U(x)$), is unique and can be expressed in the integral form as

$$\varphi(x) = -\lambda \int_{R^n} E(x,\zeta)(f(\zeta) - c)p_U(\zeta)\,d\zeta, \tag{27}$$

where $E(x,\zeta)$ is the fundamental solution of the Laplace equation.

From (27) we obtain the expression for the gradient of the potential function

$$\nabla\varphi(x) = -\lambda \int_{R^n} \nabla_x E(x,\zeta)(f(\zeta) - c)p_U(\zeta)\,d\zeta, \tag{28}$$

or

$$\nabla\varphi(x) = -\lambda E[\nabla_x E(x,U)(f(U) - c)]. \tag{29}$$

Thus, for sufficiently general assumptions on the objective function $f(x)$ concerning, in essence, only its integral properties (the correctness of the differentiation with respect to a parameter in (11)), we have obtained the representation (24), (25), (28) for the motion direction that maximizes the directional derivative of the functional $F(X)$ of the extremal problem (2) that, in turn, is equavalent to the original problem (1). This representation (also under general and natural assumptions) is unique. Therefore, the vector $\nabla\varphi$ in (28) may be considered as a substitute of the gradient for, in general, a nondifferentiable objective function f that uses only its values. The Newton potential $\varphi(x)$ in (27) may be regarded as a pattern or model of the original function $f(x)$. This model is driven by the random vector U, or, according to (19), (23), by the mean value of the function f over the distribution U and is changed with our knowledge about the minimum value of f: $\varphi = \varphi(x; c)$.

We note, in conclusion, that this Section follows, in the main, [16] and the potential function of the type (27) for global optimization was used in [28] on a heuristic basis.

6. NUMERICAL METHODS

In this section we consider briefly the numerical methods based on the potential φ (for details, see [12, 13]).

The first order methods can be written in the form

$$X_{N+1} = X_N - \epsilon_N \nabla\varphi_N(X_N), \tag{30}$$

where $\nabla\varphi_N$ is defined by (28), (29) when $c = c_N$ and $p_U = p_{U_N}$. Because of the representation of $\nabla\varphi$ in form of mathematical expectation (29), the natural way for implementation of (30) is stochastic iterative (stocha approximation) methods, which use only the realizations of the corresponding random vectors [7,22]. In this the length step should satisfy conditions

$$\sum_{N=0}^{\infty} \epsilon_N = \infty, \quad (31) \quad \sum_{N=0}^{\infty} \epsilon_N^2 < \infty.$$

The second order methods have the form

$$X_{N+1} = X_N - [\nabla^2\varphi_N(X_N)]^{-1}\nabla\varphi_N(X_N) \quad (3$$

To implement (32) we have to evaluate the inverse matrix $[\nabla^2\varphi]^{-1}$. For rather general assumptions it was shown [11], by using the notion of the singular integral, that

$$\nabla^2\varphi(x) = p(x)(\hat{f}(x)/n)I + E[\nu(U,x)]I - nE[\nu(U,x)\theta\theta^T], \quad (3$$

where $\hat{f}(x) = f(x) - c$, I is the identity matrix, $\theta = (U-x)/\|U-x\|$ is the unit random direction,

$$\nu(U,x) = \chi_{R^n \setminus S_\rho} \frac{\hat{f}(U)}{\omega_n \|U-x\|^n} + \chi_{S_\rho(U)} \frac{\hat{f}(U) - \hat{f}(x)\frac{p(x)}{p(U)}}{\omega_n \|U-x\|^n},$$

$$\chi_{S_\rho}(U) = \begin{cases} 1, & \|U-x\| \le \rho \\ 0, & \|U-x\| > \rho \end{cases} \qquad \rho > 0,$$

and ω_n is the area of the unit sphere in R^n. It can be shown [12,15] that the structure of the matrix $\nabla\varphi^2$ contai the operations of space dilatation and reflection (in perspective and nonperspective directions, respectively), whi are the base of the modern optimization methods [22] and which allow substantial simplification of the operation inversion in (32). Therefore, the Newton algorithms of the type (32) which do not use the differential characteristi of the objective function $f(x)$ must contain these operations.

The convergence analysis of the methods (30), (32) has been done in [17], where it was shown that the potenti φ can be used as a Liapunov function.

We note, in conclusion, that the second order methods can be also developed on the base of the second ord variation of the functional F in (2) [11].

7. FEASIBLE DIRECTION METHODS

As it was shown in Section 2, instead of the gradient direction Y (which maximizes $\delta_Y F(X)$), we can use directio Y, for which $\delta_Y F(X) \le 0$. In this case we find from (15) and (24), (25) that

$$\int_{R^n} (f(x) - c)\Delta\varphi(x))\, dx \;geq 0. \quad (34$$

The inequality (34) implies that

$$\begin{aligned} \Delta\varphi(x) &\ge 0, \quad x \in \Omega_N = \{x | f(x) > c\} \\ \Delta\varphi(x) &\le 0, \quad x \in \Omega_P = \{x | f(x) \le c\} \end{aligned} \quad (35)$$

Consequently, in the nonperspective domain Ω_N the potential function φ is subharmonic, while in the perspectiv domain Ω_P it is a superharmonic function [24].

In accordance with the classical Riesz theorem, any subharmonic function can be represented as a sum of harmoni and potential functions [24].

The presence of the potential φ in the structure of subharmonic functions shows that the structure of the vecto field $\bar{y}(x)$, which characterizes the motion direction, must contain the component $\nabla_x E(x,\zeta) = \omega_n \|x-\zeta\|^{-n}(x-\zeta)$ Consequently, the kernel $\nabla_x E(x,\zeta)$ is the structural (universal) component of the nonlocal search procedures tha use both gradient and, generally, feasible directions.

Let us analyze the specific features of the potential function in more details. For this we partition φ into two components,

$$\varphi(x) = \int_{\Omega_P} E(x,\zeta)\tilde{f}(\zeta)\, d\zeta + \int_{\Omega_N} E(x,\zeta)\tilde{f}(\zeta)\, d\zeta = \varphi_P(x) + \varphi_N(x), \quad (36)$$

where $\tilde{f}(\zeta) = (f(\zeta) - c)p_U(\zeta)$.

Evidently, the function $\varphi_P(x)$ is superharmonic in R^n (harmonic in Ω_N) and the function $\varphi_N(x)$ is subharmonic in R^n (harmonic in Ω_P). As a result,

$$p(x)\bar{y}(x) = \nabla\varphi_P(x) + \nabla\varphi_N(x). \tag{37}$$

If the search point x is in the nonperspective domain Ω_N, then $\delta_Y F(X) = 0$ for $\nabla\varphi_P$, because φ_P is harmonic in Ω_N ($\Delta\varphi_P(x) = 0$, for $x \in \Omega_N$). Hence, we may limit ourselves here (i.e., in Ω_N) to the component of the motion direction $\nabla\varphi_N$, where φ_N is subharmonic. According to the maximum principle for subharmonic functions [20,24], the maximum value of φ_N equals to zero (because $f(x) - c \geq 0$ and $E(x,\zeta) < 0$ for $x \in \Omega_N$) and is not achieved inside the domain Ω_N. Thus, the search motions along the direction $\nabla\varphi_N(x)$ (i.e., corresponding to the maximization of φ_N), lead to pushing the search point out of the nonperspective domain Ω_N. Since the maximization of φ_N is equivalent to finding points x for which φ vanishes, it is desirable to concentrate $p_U(x)$ at these points. This corresponds to the solution of the equation $E(x_0,\zeta)[f(\zeta) - c] = 0$, where x_0 is a realization of the current random vectorX_0 (i.e., X_0 is concentrated in the vicinity of a local minimum of f). Then the term $\|x_0 - x\|^{n-2}$ in the denominator of $E(x_0,\zeta)$ guarantees pushing of the search point out of vicinities of local minima. This is the idea of tunneling algorithms [11].

We note, in conclusion, that using the Newton potential (27) is evidently not the only way to ensure the inequality $\delta_Y F(X) \leq 0$. In particular, the wave potential $\psi(x)$, which is the solution of the wave equation

$$\Delta\psi(x) + \omega^2\psi(x) = (f(x) - c)p_U(x) \tag{38}$$

can also be used as a base for developing effective nonlocal search methods.

8. BOUNDARY CONDITIONS

At the stage of problem formulation or in the process of updating incoming information the domain of the search may be changed, that is, in the general case it is necessary to consider the problem (7) instead of (2). In this case the directional derivative of the functional F will have the form

$$\delta_Y F(X) = \int_\Omega \left(\frac{\partial f(x)}{\partial x}, p(x)\bar{y}(x)\right) dx. \tag{39}$$

Using the Stokes formula, we obtain

$$\delta_Y F(X) = \int_{\partial\Omega} (f(x) - c)(p(x)\bar{y}(x), n(x))\, ds - \int_\Omega (f(x) - c)\mathrm{div}[p(x)\bar{y}(x)]\, dx,$$

where $n(x)$ is the unit vector normal to the boundary $\partial\Omega$ of Ω. From this and the representation (24), (25) we obtain

$$\delta_Y F(X) = \int_{\partial\Omega} (f - c)(\partial\varphi/\partial n + w_n)\, ds - \int_\Omega (f - c)\Delta\varphi\, dx, \tag{40}$$

where $\partial\varphi/\partial n = (\nabla\varphi(x), n(x)); w_n = (w(x), n(x))$.

The condition of decreasing $\delta_Y F(X) \leq 0$, where $\delta_Y F(X)$ is defined in (40), leads in this case to the Neumann boundary value problem

$$\begin{array}{ll} \Delta\varphi(x) = & (f(x) - c)p_U(x), x \in \Omega \\ \partial\varphi(x)/\partial n = - & (f(x) - c)p_V(x), x \in \partial\Omega, \end{array} \tag{41}$$

where $p_U(x)$ and $p_V(x)$ are nonnegative functions which may be interpreted as probability density functions of random vectors U and V chosen on the basis of the current information in the process of problem solving.

Thus, the Neumann boundary value problem appears naturally in the process of construction of the potential field for the nonlocal search.

Instead of the Neumann boundary value problem we may also consider the Dirichlet boundary value problem

$$\begin{array}{ll} \Delta\varphi(x) = & (f(x) - c)p_U(x), x \in \Omega, \\ \varphi(x) = - & (f(x) - c)p_V(x), x \in \partial\Omega. \end{array} \tag{42}$$

However, in this case the boundary condition is in fact postulated, somewhat artificially, in contrast to the boundary value problem (41), which is derived from (40).

The solutions of the boundary value problems (41) and (42) in the integral form can be represented with Green's function, so, in general, the universal component of the nonlocal search defined by the kernel in (27) replaced by the Green's function $G(x, \zeta)$.

9. CONSIDERATION OF DIFFERENTIAL PROPERTIES OF THE OBJECTIVE FUNCTION

In the previous constructions the information about the derivatives of the objective function has not been use If we have such information we can proceed as follows. Rewrite (30) in the form

$$\nabla\varphi(x) = -\lambda \int_{R^n} \nabla_\zeta E(x, \zeta)(f(\zeta) - c)p_U(\zeta)\, d\zeta. \qquad (4$$

Applying Green's formula to (43) we obtain

$$\nabla\varphi(x) = -\lambda \int_{R^n} \nabla[f(\zeta) - c)p_U(\zeta)]E(x, \zeta)\, d\zeta,$$

or, finally,

$$\nabla\varphi(x) = -\lambda \int_{R^n} E(x, \zeta)\nabla f(\zeta)p_U(\zeta)\, d\zeta - \qquad (4$$

$$-\lambda \int_{R^n} E(x, \zeta)(f(\zeta) - c)\nabla p_U(\zeta)\, d\zeta.$$

Thus, the local information on ∇f (the first term in the (44)) is added to the second "nonlocal" term in (44).

10. STRUCTURAL COMPONENTS OF THE NONLOCAL SEARCH

Let us review the chain of assumptions and constructions that has led us to the definition of the potential ar to the algorithms for its minimization. First of all, we note that in the case considered the transition from th original to the randomized problem is the equivalent operation (in the sense of asymptotic solutions), but it open new possibilities for developing nonlocal search procedures. The subsequent constructions also were made und natural and nonrestrictive assumptions (such as the existence of integrals or the validity of their differentiation wit respect to a parameter). Therefore, we may speak of the structural (universal) components that must be found i nonlocal search procedures.

Indeed, analysis of the constructions (1) – (30) that has led us to the definition of the potential shows that th kernel $E(x, \zeta)$ (or, generally, the Green's function $G(x, \zeta)$)) is such a universal component of nonlocal search tha depends neither on the way of randomization, nor on the choice of the objective function f. Analysis of the firs order derivatives of this kernel shows that nonlocal search must have an unstable component in the nonperspectiv domain, while analysis of the second order derivatives leads (also with necessity) to the space dilatation-reflectio operations in the nonlocal search methods.

In conclusion we dwell briefly on the role of the "nondivergent" vector field $w(x)$ in the nonlocal search. Sub stituting (24), (25) into (18), (22) and assuming that $\Delta\varphi(x) = 0$, we obtain $\delta_Y F(X) = 0$. Thus, the motion whic is determined by the vector field $w(x)$ does not change the mean value of the function f and is necessary for th accumulation of our knowledge about the f. Therefore, the uncertainty in the representation (24), (25) associate with the choice of the vector field $w(x)$ may be considered as a "penalty" for the absence of a priori informatio about the properties of the objective function f.

11. CONCLUSION

In the present paper we have concentrated on the theoretical aspects of the research direction considered. Th numerical methods are discussed in details in [12, 13]. Here we only note the succession of these algorithms (o the level of their basic operations or structural components) from the known deterministic analogs (methods of th first and the second order, the conjugate directions method, the space dilatation methods, the ellipsoidal methods etc). We should also emphasize that in the algorithms based on the potential function only the values of the origina objective function are used.

Finally, we note that in this paper we have used a potential depending on the integral properties of the objectiv function $\varphi = \varphi(x, c), c = \int f(x)p_U(x)\, dx$. It is also possible to use a potential function that depends on the averaged solution (the approximation "center") $\varphi = \varphi(x, \bar{x}), \bar{x} = \int x p_U(x)\, dx$. In this case the corresponding methods wil

be similar to local approximation methods [27] or to nonparametric density estimation methods [28] and are briefly discussed in [29].

REFERENCES

1. S. Fromovitz, "Nonlinear programming with randomization," *Management Science*, **11**, No. 9, 831–846 (1965).
2. A. I. Kaplinskii and A. I. Propoi, "On stochastic approach to nonlinear programming problems", *Avtomat. Telemekh.*, No. 3, 122–133 (1970).
3. Yu. M. Ermol'ev, "A general problem on stochastic programming" [in Russian], *Kibernetika*, No. 3, 47–50 (1971).
4. D. B. Yudin, *Mathematical Methods of Control Under Conditions of Uncertainty* [in Russian], Sovetskoe Radio, Moscow (1974).
5. A. I. Kaplinskii, A. S. Poznyak, and A. I. Propoi, "Optimality conditions for certain stochastic programming problems," *Avtomat. Telemekh.*, No. 8, 51–60 (1971).
6. A. I. Kaplinskii, A. S. Poznyak, and A. I. Propoi, "Methods for solutions of stochastic programming problems," em Avtomat. Telemekh., No. 10, 87–94 (1971).
7. Ja. Z. Tsypkin, *Adaptation and Learning in Automatic Systems* [in Russian], Nauka, Moscow (1968) (English: Academic Press, New York (1971)).
8. M. A. Aizerman, E. M. Braverman, and L. I. Rozonoer, *Methods of Potential Functions in the Theory of Machine Learning* [in Russian], Nauka, Moscow (1970).
9. A. I. Propoi and A. V. Pukhlikov, *Foundations of Mathematical Systems Theory* [in Russian], VNIISI, Moscow (1990).
10. A. I. Kaplinskii, E. E. Limarev, and G. D. Chernysheva, "Design of randomized optimization algorithms" [in Russian], *Problemy Sluchainogo Poiska*, No. 8, 63–91 (1980).
11. A. I. Kaplinskii and A. I. Propoi, *Variational Approach to Construction of Nonlocal Optimization Algorithms* [in Russian], VNIISI, Moscow (1986).
12. A. I. Kaplinskii and A. I. Propoi, *Structural Components of Nonlocal Search Using Potential Theory* [in Russian], VNIISI, Moscow (1990).
13. A. I. Kaplinskii and A.I. Propoi, *Development of Numerical Methods for Nonlocal Search Using Potential Theory* [in Russian], VNIISI, Moscow (1990).
14. A. B. Gil', A. I. Kaplinskii, and A. I. Propoi, "Design of numerical schemes for nonlocal optimization on the basis of potential theory," in: *Statistical Methods and Models* [in Russian], VNIISI, Moscow, 72–78 (1987).
15. A. B. Gil', A. I. Kaplinskii, and A. I. Propoi, "On the use of potential theory in the variational approach to design of optimization algorithms using reflection and dilatation operations," in: *Models and Methods of Optimization* [in Russian], VNIISI, Moscow, 11–23 (1987).
16. A. I. Kaplinskii and A. I. Propoi, "The gradient basis for nonsmooth optimization using potential theory," in: *Problems and Methods of Optimization Modelling* [in Russian], VNIISI, Moscow, 11–16 (1988).
17. A. I. Kaplinskii, A. M. Pesin, and A. I. Propoi, "Methods of nonlocal search," in: *Models and Methods of Optimization* [in Russian], VNIISI, Moscow, 35–45 (1991).
18. Yu. V. Egorov, *Linear Differential Equations of Principal Type* [in Russian], Nauka, Moscow (1984).
19. S. Zacks, *The Theory of Statistical Inference*, John Willey, New York, (1971).
20. A. N. Tikhonov and A. A. Samarskii, *The Equations of Mathematical Physics* [in Russian], Nauka, Moscow (1972).

21. A. A. Krasovskii, "Continuous algorithms and stochastic dynamics of search for extremum," *Avtomat. Te mekh.*, No. 4, 55–65 (1991).

22. M. B. Nevel'son and R. Z. Has'minskii, *Stochastic Approximation and Reccurrent Estimation* [in Russia Nauka, Moscow (1972).

23. N. Z. Shor, *Minimization Methods for Nondifferential Functions and Their Applications* [in Russian], Nauko Dumka, Kiev (1979).

24. W. K. Hayman and P. B. Kennedy, *Subharmonic Functions*, Academic Pres, New York (1976).

25. A. V. Levy and A. Montalvo, "The tunneling algorithm for the global minimization of functions", *SIAM Sci. Comp.*, 6, No. 1, 15–29 (1985).

26. E. Sanchez-Palencia, *Non-Homogenous Media and Vibration Theory*, Springer-Verlag, New York (1985).

27. V. Ya. Katkovnik, *Linear Estimates and Stochastic Optimization Problems* [in Russian], Nauka, Moscow (1976

28. L. Devroye and L. Gyorfi, *Nonparametric Density Estimation*, John Willey, New York (1985).

29. A. I. Propoi, "On principles of nonlocal search," in: *Optimization and Control in Complex Systems* [in Russian VNIISI, Moscow, 67–80 (1989).